

Dörfler, Tobias; Golke, Stefanie; Artelt, Cordula

**Dynamisches Testen der Lesekompetenz. Theoretische Grundlagen, Konzeption und Testentwicklung. Projekt Dynamisches Testen**

*Klieme, Eckhard [Hrsg.]; Leutner, Detlev [Hrsg.]; Kenk, Martina [Hrsg.]: Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. Weinheim ; Basel : Beltz 2010, S. 154-164. - (Zeitschrift für Pädagogik, Beiheft; 56)*



**Quellenangabe/ Reference:**

Dörfler, Tobias; Golke, Stefanie; Artelt, Cordula: Dynamisches Testen der Lesekompetenz. Theoretische Grundlagen, Konzeption und Testentwicklung. Projekt Dynamisches Testen - In: Klieme, Eckhard [Hrsg.]; Leutner, Detlev [Hrsg.]; Kenk, Martina [Hrsg.]: Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. Weinheim ; Basel : Beltz 2010, S. 154-164 - URN: urn:nbn:de:0111-opus-34041 - DOI: 10.25656/01:3404

<https://nbn-resolving.org/urn:nbn:de:0111-opus-34041>

<https://doi.org/10.25656/01:3404>

in Kooperation mit / in cooperation with:

**BELTZ**

<http://www.beltz.de>

**Nutzungsbedingungen**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.  
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

**Terms of use**

We grant a non-exclusive, non-transferable, individual and limited right to using this document.  
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

**Kontakt / Contact:**

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Zeitschrift für Pädagogik · 56. Beiheft

# **Kompetenzmodellierung**

## **Zwischenbilanz des DFG- Schwerpunktprogramms und Perspektiven des Forschungsansatzes**

Herausgegeben von

Eckhard Klieme, Detlev Leutner und Martina Kenk

**BELTZ**

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder ein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder auf ähnlichem Wege bleiben vorbehalten. Fotokopien für den persönlichen oder sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopie hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte oder genutzte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG Wort, Abteilung Wissenschaft, Goethestr. 49, 80336 München, bei der die einzelnen Zahlungsmodalitäten zu erfragen sind.

© 2010 Beltz Verlag · Weinheim und Basel

Herstellung: Lore Amann

Gesamtherstellung: Druckhaus „Thomas Müntzer“, Bad Langensalza

Printed in Germany

ISSN 0514-2717

Bestell-Nr. 41157

# Inhaltsverzeichnis

*Eckhard Klieme/Detlev Leutner/Martina Kenk*

Kompetenzmodellierung. Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms. Einleitung zum Beiheft .....	9
--	---

*Benő Csapó*

Goals of Learning and the Organization of Knowledge .....	12
---	----

## Mathematische Kompetenzen

*Marianne Bayrhuber/Timo Leuders/Regina Bruder/Markus Wirtz*

Projekt HEUREKO

Repräsentationswechsel beim Umgang mit Funktionen – Identifikation von Kompetenzprofilen auf der Basis eines Kompetenzstrukturmodells .....	28
---	----

*Andreas Frey/Nicki-Nils Seitz*

Projekt MAT

Multidimensionale adaptive Kompetenzdiagnostik: Ergebnisse zur Messeffizienz .....	40
--	----

*Nina Zeuch/Hanneke Geerlings/Heinz Holling/Wim J. van der Linden/*

*Jonas P. Bertling*

Projekt Regelgeleitete Itementwicklung

Regelgeleitete Konstruktion von statistischen Textaufgaben: Anwendung von linear logistischen Testmodellen und Aufgabencloning .....	52
--	----

*Eckhard Klieme/Anika Bürgermeister/Birgit Harks/Werner Blum/Dominik Leiß/*

*Katrin Rakoczy*

Projekt Co<sup>2</sup>CA

Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht .....	64
--	----

*Olga Kunina-Habenicht/Oliver Wilhelm/Franziska Matthes/André A. Rupp*

Projekt Kognitive Diagnosemodelle

Kognitive Diagnosemodelle: Theoretisches Potential und methodische Probleme ...	75
---	----

*Aiso Heinze*

Review

Mathematische Kompetenz modellieren und diagnostizieren: Eine Diskussion der Forschungsprojekte des DFG-Schwerpunktprogramms „Kompetenzmodelle“ aus mathematikdidaktischer Sicht .....	86
--	----

## Naturwissenschaftliche Kompetenzen

*Tobias Viering/Hans E. Fischer/Knut Neumann*

Projekt Physikalische Kompetenz

Die Entwicklung physikalischer Kompetenz in der Sekundarstufe I .....	92
---	----

*Renate Soellner/Stefan Huber/Norbert Lenartz/Georg Rudinger*

Projekt Gesundheitskompetenz

Facetten der Gesundheitskompetenz – eine Expertenbefragung .....	104
--	-----

*Ilonca Hardy/Thilo Kleickmann/Susanne Koerber/Daniela Mayer/*

*Kornelia Möller/Judith Pollmeier/Knut Schwippert/Beate Sodian*

Projekt Science – P

Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter .....	115
---	-----

*Nina Roczen/Florian G. Kaiser/Franz X. Bogner*

Projekt Umweltkompetenz

Umweltkompetenz – Modellierung, Entwicklung und Förderung .....	126
---	-----

*Ilka Parchmann*

Review

Kompetenzmodellierung in den Naturwissenschaften – Vielfalt ist wertvoll, aber nicht ohne ein gemeinsames Fundament .....	135
---	-----

## Sprachliche und Lesekompetenzen

*Wolfgang Schnotz/Nele McElvany/Holger Horz/Sascha Schroeder/Mark Ullrich/*

*Jürgen Baumert/Axinja Hachfeld/Tobias Richter*

Projekt BITE

Das BITE-Projekt: Integrative Verarbeitung von Bildern und Texten in der Sekundarstufe I .....	143
--	-----

*Tobias Dörfler/Stefanie Golke/Cordula Artelt*

Projekt Dynamisches Testen

Dynamisches Testen der Lesekompetenz: Theoretische Grundlagen, Konzeption und Testentwicklung .....	154
---	-----

*Thorsten Roick/Petra Stanat/Oliver Dickhäuser/Volker Frederking/  
Christel Meier/Lydia Steinhauer*

Projekt Literarästhetische Urteilskompetenz

Strukturelle und kriteriale Validität der literarästhetischen Urteilskompetenz ..... 165

*Hans Anand Pant/Simon P. Tiffin-Richards/Olaf Köller*

Projekt Standard-Setting

Standard-Setting für Kompetenztests im Large-Scale-Assessment ..... 175

*Johannes Hartig/Jana Höhler*

Projekt MIRT

Modellierung von Kompetenzen mit mehrdimensionalen IRT-Modellen ..... 189

*Albert Bremerich-Vos*

Review

Modellierung von Aspekten sprachlich-kultureller Kompetenz. Anmerkungen  
zu den Projektberichten ..... 199

## **Fächerübergreifende Kompetenzen**

*Ellen Gausmann/Sabina Eggert/Marcus Hasselhorn/Rainer Watermann/  
Susanne Bögeholz*

Projekt Bewertungskompetenz

Wie verarbeiten Schüler/-innen Sachinformationen in Problem- und  
Entscheidungssituationen Nachhaltiger Entwicklung – Ein Beitrag zur

Bewertungskompetenz ..... 204

*Samuel Greiff/Joachim Funke*

Projekt Dynamisches Problemlösen

Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal  
komplexer Systeme ..... 216

*Klaus Lingel/Nora Neuenhaus/Cordula Artelt/Wolfgang Schneider*

Projekt EWIKO

Metakognitives Wissen in der Sekundarstufe: Konstruktion und Evaluation  
domänenspezifischer Messverfahren ..... 228

*Jens Fleischer/Joachim Wirth/Stefan Rumann/Detlev Leutner*

Projekt Problemlösen

Strukturen fächerübergreifender und fachlicher Problemlösekompetenz –  
Analyse von Aufgabenprofilen ..... 239

*Melanie Schütte/Joachim Wirth/Detlev Leutner*

Projekt Selbstregulationskompetenz

Selbstregulationskompetenz beim Lernen aus Sachtexten – Entwicklung und  
Evaluation eines Kompetenzstrukturmodells ..... 249

*Tobias Gschwendtner/Bernd Geißel/Reinhold Nickolaus*

Projekt Berufspädagogik

Modellierung beruflicher Fachkompetenz in der gewerblich-technischen  
Grundbildung ..... 258

*Franziska Perels*

Review

Modellierung und Messung fächerübergreifender Kompetenzen und ihre  
Bedeutung für die Bildungsforschung. Kritische Reflexion der Projektbeiträge ... 270

## **Lehrerkompetenzen**

*Simone Bruder/Julia Klug/Silke Hertel/Bernhard Schmitz*

Projekt Beratungskompetenz

Modellierung der Beratungskompetenz von Lehrkräften ..... 274

*Cornelia Gräsel/Sabine Krolak-Schwerdt/Ines Nölle/Thomas Hörstermann*

Projekt Diagnostische Kompetenz

Diagnostische Kompetenz von Grundschullehrkräften bei der Erstellung der  
Übergangsempfehlung: eine Analyse aus der Perspektive der sozialen  
Urteilsbildung ..... 286

*Tina Seidel/Geraldine Blomberg/Kathleen Stürmer*

Projekt OBSERVE

„OBSERVER“ – Validierung eines videobasierten Instruments zur Erfassung  
der professionellen Wahrnehmung von Unterricht ..... 296

*Mareike Kunter*

Review

Modellierung von Lehrerkompetenzen. Kommentierung der  
Projektdarstellungen ..... 307

# Dynamisches Testen der Lesekompetenz

*Theoretische Grundlagen, Konzeption und Testentwicklung*

*Projekt Dynamisches Testen<sup>1</sup>*

## 1. Einleitung

Dynamische Tests gelten insbesondere im Bildungssektor als Alternative zu herkömmlichen Tests. Für den Bereich der Lesekompetenz fehlen bislang jedoch gesicherte Erkenntnisse, die die diagnostische Güte eines dynamischen Lesekompetenztests belegen können. Der hier vorliegende Beitrag bietet einen theoretischen Überbau für die Entwicklung eines solchen dynamischen Lesekompetenztests und informiert zudem (nicht immer chronologisch) über grundlegende Schritte der Testentwicklung sowie erste Ergebnisse der Testkonstruktion.

## 2. Theoretischer Rahmen

### *2.1 Feedback und (meta-) kognitive Hilfen als Elemente dynamischer Tests*

Kognitive Fähigkeiten sind nicht direkt beobachtbar, sondern müssen aus Verhaltensindikatoren erschlossen werden. Hierzu werden Personen in der Regel mit einer Reihe von Testaufgaben konfrontiert, deren Bearbeitung unter Berücksichtigung eines Messfehlers Aufschluss über die Ausprägung ihrer zugrunde liegenden Fähigkeit(en) geben soll. Beachtenswert ist, dass sich auf diese Weise nur derjenige Anteil der Leistungsfähigkeit abbilden lässt, der bereits in ausreichendem Maße im Verhalten der Proband/innen realisiert wurde. Aussagen über mögliche Leistungsreserven können dadurch nicht getroffen werden. Ist neben der Diagnostik der aktuellen Leistung zusätzlich die Abschätzung von individuellen Leistungsreserven bzw. einer Lernfähigkeit im untersuchten Bereich von Interesse, kommen dynamische Tests zum Einsatz (vgl. Guthke/Beckmann/Wiedl 2003).

Die Kernidee zur Erfassung der Lernfähigkeit geht auf Wygotski (1964) zurück, der konstatiert, dass eine zuverlässige Diagnose nur unter Beachtung zweier sogenannter Entwicklungszonen erfolgen kann. Nach seiner Auffassung muss für die Bewertung der Leistungsfähigkeit einer Person neben dem aktuellen Entwicklungsstand („Zone der ak-

---

1 Diese Veröffentlichung wurde ermöglicht durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft (Kennz.: AR 301/7-2) im Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (SPP 1293).



tuellen Entwicklung“) zusätzlich ihre Entwicklungspotenz („Zone der nächsten Entwicklung“) erfasst werden; also die Fähigkeit eines Individuums, unter förderlichen Bedingungen (z.B. durch Hilfe von Erwachsenen, fähigeren Peers oder computergestützten tutoriellen Systemen) Leistungsverbesserungen zu erzielen. Diese Entwicklungspotenzen können auch durch einen simulierten Lernvorgang im Testprozess – der durch gezielte Feedbacks und spezifische Lern- und Denkhilfen von außen gesteuert wird – „sichtbar“ gemacht werden. Das Ausmaß, in dem ein/e Proband/in von den angebotenen Feedbacks und Hilfen profitiert, gilt als Indikator der intellektuellen Lernfähigkeit.

Feedback kann in diesem Zusammenhang als Informationsangebot an eine Person bezüglich ihrer Leistung oder Verstehensprozesse beschrieben werden, mit dem Ziel, regulierend auf zukünftige Prozesse einzuwirken (vgl. Hattie/Timperley 2007). Es bietet mit einer zumeist mittleren Effektstärke eine erfolgversprechende Möglichkeit zur Leistungsverbesserung (vgl. Bangert-Drowns u.a. 1991; Kluger/DeNisi 1996). Die Qualität der rückgemeldeten Information reicht nach Narciss (2006) von einfachen bis hin zu sogenannten elaborierten Feedback-Arten.

Elaborierte Feedback-Arten, die beispielsweise fehlerspezifische Korrekturhinweise oder Hinweise auf kognitive bzw. metakognitive Strategien bieten, sind – unabhängig vom untersuchten Kompetenzbereich – zur Leistungsverbesserung besonders geeignet (Shute 2008). Während kognitive Strategien bei einfachen und komplexen Anforderungen gleichermaßen bedeutsam sind, gewinnen metakognitive Strategien mit zunehmender Komplexität der Aufgabe an Bedeutung. Als Schlüssel hierfür gelten übergeordnete Strategien der Planung, Überwachung und Bewertung und die darauf basierende Regulation des eigenen Lernvorgangs. Zentrale Momente metakognitiver Aktivitäten sind zum einen die Reflexion über den eigenen Lernprozess und zum anderen sind es die durch diese Reflexionen ausgelösten strategischen Aktivitäten. Die sichere Anwendung lösungsrelevanter kognitiver und metakognitiver Strategien ist beim verstehenden Lesen von besonderer Relevanz (vgl. Artelt/Baumert/Julius-McElvany 2003).

## 2.2 Lesekompetenz/Textverstehen als Anwendungsfeld dynamischer Tests

Das Konzept des Dynamischen Testens ist im Bereich der Intelligenzdiagnostik am weitesten verbreitet, es kann jedoch nach Guthke, Beckmann und Wiedl (2003) auch auf andere Kompetenzbereiche übertragen werden. Lesen umfasst kognitive Anforderungen, für die ähnliche Annahmen getroffen werden können wie für die verstärkt im Rahmen dynamischer Tests untersuchten Facetten der Intelligenz: Das Verstehen von Texten ist wie das Lösen von Aufgaben in einem Intelligenztest eine komplexe kognitive Anforderung, die durch Hilfen und Rückmeldungen unterstützt werden kann.

Textverstehen ist ein partiell automatisch ablaufender, aktiver Interaktionsprozess zwischen Leser/in und Text. Dabei werden, teilweise unter Nutzung des vorhandenen Vorwissens, verschiedene kognitive Repräsentationen des Textes aufgebaut, die nicht unabhängig voneinander sind (vgl. Kintsch 1998): Bei der wörtlichen Repräsentation stehen Oberflächenmerkmale des Textes im Vordergrund, die ihrerseits eine wesentliche

Voraussetzung für die semantische und syntaktische Analyse der Textinhalte bilden. Wird diese (textbasierte) propositionale Repräsentation des Geschriebenen durch zusätzliche (textferne) Propositionen angereichert, entsteht eine situative Repräsentation des Textes, ein sogenanntes Situationsmodell. Die Anreicherung mit Propositionen, die keine direkte Entsprechung im Text haben, geschieht unterdessen über inferenzielle und vorwissensabhängige Prozesse (vgl. Richter/Christmann 2002).

Inferenzen werden allerdings bereits auf hierarchieniedrigen Repräsentationsebenen beim Textverstehen gebildet. Hierzu zählen beispielsweise lokale Kohärenzbildungen im Sinne von identischen, synonymen oder abstrakten Wortwiederholungen. Zu den hierarchiehoher Verarbeitungsprozessen beim Lesen wird z.B. der Prozess der globalen Kohärenzbildung gezählt, der u.a. durch thematische und elaborative Inferenzen gekennzeichnet ist (vgl. Graesser/Singer/Trabasso 1997). Die erfolgreiche Generierung von Inferenzen ist laut Oakhill und Garnham (1988) ein anspruchsvoller – und im Falle hierarchiehoher Prozesse auch kapazitätsintensiver – intellektueller Prozess, der eine wesentliche Bedingung für verstehendes Lesen darstellt: Gute Leser/innen unterscheiden sich von schlechten Lesern/Leserinnen u.a. auch durch die Anzahl der beim Lesen gezogenen Inferenzen (vgl. ebd.; Oakhill/Cain 2004).

Ein tieferes Verstehen eines Textes, der nicht allein auf leicht verfügbarem Weltwissen beruht, bedarf zusätzlich der intentionalen und strategischen Steuerung des Lesevorgangs (vgl. Coté/Goldman 1999). Der Aufbau einer kohärenten Textrepräsentation hängt – besonders bei längeren und schwierigen Texten – stark davon ab, ob und in welchem Maße kognitive und metakognitive Strategien effektiv verwendet werden. Hierzu zählt etwa die kontinuierliche Überprüfung des Verständnisses des gelesenen Textes und die entsprechende Regulation des Leseprozesses, z.B. durch das erneute Lesen einzelner Textstellen oder den gezielten Abruf von Informationen aus dem Langzeitgedächtnis (vgl. Kintsch 1998). Die nicht prinzipiell automatisch ablaufenden Prozesse beim verstehenden Lesen werden unter dem Strategiebegriff subsumiert. Die Unterstützung der Leserin/des Lesers bei der strategischen Bearbeitung eines Textes ist daher Ziel zahlreicher Trainingsprogramme (vgl. z.B. Souvignier/Mokhesgerami 2006).

Bei der Konstruktion eines dynamischen Lesekompetenztests sind die Erkenntnisse aus der Feedback- und Trainingsliteratur ebenso zu berücksichtigen wie die Besonderheiten der Entwicklung von Testaufgaben in diesem Bereich. Der folgende Abschnitt liefert eine Zusammenfassung zur theoretischen Konzeption eines solchen dynamischen Lesekompetenztests.

### 3. Konzeption eines dynamischen Lesekompetenztests

Bei einem dynamischen Test der Lesekompetenz müssen Testaufgaben zwei spezifische Anforderungen erfüllen: Einerseits müssen sie Prozesse des Textverstehens abbilden, um eine valide Diagnose der Textverstehensleistung zu gewährleisten; andererseits müssen die Aufgaben zusätzlich genügend Möglichkeiten zur Implementierung von passenden Feedbacks und Hilfen bieten.

Um beiden Anforderungsbereichen im hier vorgestellten Projekt gerecht zu werden, wurden Texte und Aufgaben teils adaptiert, teils neu konstruiert. Die Konstruktion und Auswahl von Aufgaben zur Messung des Textverstehens erfolgte angelehnt an die o.g. Unterscheidung nach hierarchieniedrigen und -hohen Anforderungen und dabei insbesondere an inferenziellen Prozessen (vgl. Kintsch 1998), die im Folgenden näher erläutert werden. Im Test werden Indikatoren lokaler Kohärenzbildung (Herstellung eines Sinnzusammenhangs zwischen aufeinanderfolgenden Propositionen oder Sätzen), Indikatoren globaler Kohärenzbildung (Herstellung eines Sinnzusammenhangs zwischen größeren Textteilen auf höherer Abstraktionsebene) und Indikatoren aktiv-konstruktiver Prozesse zur Bildung einer inhaltspezifischen und anschaulichen Repräsentation des im Text beschriebenen Sachverhalts (Situationsmodell) abgebildet.

Bezogen auf diese Textverstehensprozesse orientierte sich die Konstruktion der Feedbacks und (meta-)kognitiven Hilfen an zwei Forschungsrichtungen. Einerseits wurden Erkenntnisse aus dem Bereich der Feedbackforschung, die sich mit Auswirkungen gezielter Rückmeldungen auf die Testleistung beschäftigt (vgl. u.a. Bangert-Drowns u.a. 1991; Hattie/Timperley 2007; Kluger/DeNisi 1996; Shute 2008), genutzt. Feedback scheint demnach dann am wirksamsten zu sein, wenn es spezifische Fehler korrigiert bzw. verständnisfördernd wirkt. Andererseits musste die spezifische Domäne – Lesekompetenz/Textverstehen – bei der Konstruktion von Rückmeldungen beachtet werden. Um eine Verständnisförderung beim Lesen durch Feedbacks bzw. Hilfen zu erreichen, sollten jene inferenziellen und strategischen Prozesse unterstützt werden, die für die Bildung einer kohärenten Textrepräsentation bezogen auf die jeweiligen Aufgabentypen bedeutsam sind (vgl. z.B. Graesser/Singer/Trabasso 1997). Die notwendigen Erkenntnisse zur Wirksamkeit unterschiedlicher Feedbackarten für die konstruierten Textverstehensanforderungen konnten bereits in einer experimentellen Untersuchung gewonnen werden (vgl. Golke/Dörfler/Artelt 2009).

### 3.1 Testentwicklung

Die bislang geleistete Testentwicklung diente primär dem Ziel, für das o.g. Experiment einen angemessenen Lesekompetenztest zur Verfügung zu stellen. Da diese vorläufige Testkonstruktion auch zur Überprüfung der Aufgabenqualität im Sinne einer Pilotstudie herangezogen werden kann, wird auf die Ergebnisse der ersten Phase der Testentwicklung nachfolgend genauer eingegangen.

Nach der Auswahl von Textmaterial und der Auswahl und Konstruktion entsprechender Items (s.u.) wurden zunächst kognitive Interviews mit Schüler/innen der 6. Klassenstufe durchgeführt, mit dem Ziel, Texte und Aufgaben hinsichtlich ihrer Angemessenheit für diesen Altersbereich zu überprüfen und mögliche Verständnisschwierigkeiten zu eruieren. Insgesamt wurden dabei 20 Schüler/innen mit Texten und Aufgaben konfrontiert und nach der Wirksamkeit eingesetzter Feedbacks befragt. Die Interviews dauerten etwa 1 Stunde. Hierbei wurden Fragen zum allgemeinen Verständnis des Textes und der Aufgaben gestellt und die Schüler/innen aufgefordert, laut bei der Bearbei-

### Durch die Eiswüste der Antarktis

Noch um 1900 war es keinem Forscher gelungen, zum Nordpol oder Südpol vorzudringen; zahlreiche Expeditionen scheiterten. Erst im Jahr 1909 erreicht der Amerikaner Peary mit Schlittenhunden den Nordpol. Der Norweger Amundsen, zur selben Zeit mit einer Expedition zum Nordpol unterwegs, erhält die Nachricht von Pearys Erfolg. Daraufhin ändert er seinen Plan und entschließt sich zu einer neuen Expedition Richtung Südpol: „Ich werde in den Süden gehen!“. Doch er ahnt noch nicht, dass er da nicht der Einzige ist.

Das gleiche Ziel hat nämlich auch der Engländer Scott. Ein verbissener „Wettlauf“ entsteht. Wer wird als Erster am Südpol stehen? Im Januar 1911 erreichen beide Expeditionen die die Antarktis und errichten, 600 km

Das ist falsch. Die Vorratslager werden von Amundsens und Scotts Mannschaft selbst angelegt, das heißt die Vorratslager werden nicht verteilt.

#### Frage 2

Warum entsteht zwischen Amundsen und Scott ein „verbissener Wettlauf“?

- ☒ Wer als Erster die Antarktis erreicht, bekommt die meisten Vorratslager.
- ☐ Wer als Erster die Antarktis erreicht, ist der allererste Forscher am Südpol.
- ☐ Die Strapazen der Expedition sind im Wettkampf besser zu ertragen.
- ☐ Wer als Letzter in der Antarktis zurückbleibt, kommt dort zu Tode.
- ☐ Beide beeilen sich, um den Südpol vor Einbruch des Winters zu erreichen.

WEITER

Abb. 1: Beispielhafte Darstellung einer Lesekompetenzaufgabe

tung der Materialien zu denken. Die Durchführung der kognitiven Interviews erfolgte in Anlehnung an Lewis und Reiman (1993) sowie Willis (2004).

Für die Testentwicklung wurden drei expositorische und zwei narrative Texte mit insgesamt 37 Aufgaben verwendet. Es wurde ein computerbasiertes Multiple-Choice-Format mit fünf Antwortalternativen umgesetzt. Das Testformat wird in Abbildung 1 schematisch dargestellt. Während der Bearbeitung der Aufgaben war der jeweilige Text stets verfügbar. Vor- und Zurückblättern zwischen Texten und Aufgaben war nicht möglich.

### 3.1.1 Stichprobe

Die neu entwickelten Texte und Aufgaben wurden insgesamt 566 bayerischen Schüler/innen und Schülern vorgelegt. Die Schüler/innen verteilten sich hinreichend gleichmä-

Big auf die drei Schulformen Haupt-, Realschule und Gymnasium. Beide Geschlechter waren zu gleichen Anteilen in der Stichprobe vertreten.

### Skalierung

In den folgenden Abschnitten werden erste Ergebnisse der Testentwicklung berichtet. Hierbei werden insbesondere teststatistische Kennwerte dargestellt, die über die Güte des Testprototypens informieren sollen.

*Itemparameter.* Aus der bisherigen Testentwicklung resultierten 37 Aufgaben. Bei der Analyse der Items wird angenommen, dass alle Anforderungen eine zugrunde liegende Fähigkeitsdimension (Lesekompetenz) abbilden. Zur Modellierung der Antwortwahrscheinlichkeiten wird daher ein eindimensionales Raschmodell mit der Software ConQuest (vgl. Wu/Adams/Wilson 1997) geschätzt. Zur Bewertung der Items werden Itemfit-Kennwerte (Weighted Mean Square Residualwerte (MNSQ), *T*-Werte), Itemtrennschärfeparameter und Itemschwierigkeiten herangezogen. Zudem können erwartete und beobachtete itemcharakteristische Kurven (ICC) verglichen werden, deren gleichförmiger Verlauf einen guten Indikator für die Modellpassung in allen untersuchten Fähigkeitsbereichen liefert. Die Reliabilität der Gesamtskala liegt bei  $\alpha = .76$ .

Die Items liefern insgesamt zufriedenstellende itemstatistische Kennwerte (Tabelle 1). Lediglich vier Items zeigen einen leicht auffälligen gewichteten MNSQ-Wert. Sowohl die Verläufe der ICCs der vier Items als auch die herkömmlichen Itemschwierigkeits- und Trennschärfeparameter sind jedoch unauffällig. Drei weitere Aufgaben zeigen bei unbedenklichen Fitindizes unterdurchschnittlich niedrige Trennschärfen.

Anhand der gemeinsamen Verteilung der Personen- und Aufgabenparameter kann festgestellt werden, dass zahlreiche Aufgaben für die untersuchten Sechstklässler/innen tendenziell zu schwierig waren. Dieser unter herkömmlichen Testbedingungen kritische Umstand ist jedoch für weitere Projektschritte notwendig: Um Feedbacks in den Testprozess zu implementieren, sind tendenziell zu schwierige Items besser geeignet als Items mit optimaler Schwierigkeitspassung. Ziel der Rückmeldungen im Zuge dynamischer Tests ist es schließlich, Leistungsreserven abzubilden. Dazu ist es notwendig, intellektuell fordernde Aufgaben anzubieten, um den Testpersonen die Möglichkeit zur Leistungssteigerung einzuräumen.

*Differenzielle Itemfunktionen.* Die bislang entwickelten Aufgaben sind mit der Annahme eines eindimensionalen Lesekompetenztests vereinbar. Als weitere Voraussetzung für die Güte des konstruierten Tests wurde geprüft, ob die Aufgaben wie intendiert alle in gleicher Weise Fähigkeitsunterschiede zwischen Personen abbilden, oder ob in Abhängigkeit vom Geschlecht oder der Schulform über die Unterschiede im Gesamttest hinausgehende Vor- oder Nachteile bei einzelnen Aufgaben existieren (sog. Differenzielle Itemfunktionen, DIF). Die DIF-Analyse nach Geschlecht ergab, dass es bei einem Item deutliche und über die im Gesamttest zu findenden Unterschiede (s.u.) zwischen Jungen und Mädchen hinausgehende Vorteile zugunsten der Mädchen gab ( $-0.53$  logits), was auf ein aus diagnostischer Perspektive schlechtes, da unfaires Item hindeutet. Die DIF-Analysen hinsichtlich der erfassten Schulformen ergeben beim Vergleich der beiden Extreme Hauptschule und Gymnasium bei sechs der 37 Aufgaben einen über die

Item	Estimate	MSNQ	$T$	$p_i$	$r_{it}$
1	0.61	1.03	0.7	27.48	0.26
2	-0.76	1.01	0.5	57.02	0.32
3	0.73	1.03	0.5	25.40	0.25
4	0.44	1.10	2.4	30.73	0.13
5	0.09	0.98	-0.7	38.08	0.38
6	0.29	1.08	2.1	33.81	0.18
7	-0.05	0.94	-2.3	41.13	0.44
8	1.25	1.10	1.3	17.20	0.04
9	-0.24	0.96	-1.4	45.21	0.40
10	-0.22	0.95	-2.0	44.86	0.43
11	0.62	0.96	-0.8	27.30	0.38
12	0.23	1.09	2.5	35.11	0.18
13	0.57	1.04	1.0	28.19	0.23
14	-0.64	0.95	-1.9	54.43	0.43
15	-0.49	0.98	-0.9	50.89	0.39
16	0.55	1.06	1.5	28.72	0.19
17	-0.66	0.94	-2.5	54.96	0.46
18	-0.06	1.09	3.1	41.31	0.18
19	0.59	1.10	2.2	27.84	0.12
20	-0.63	0.94	-2.5	54.08	0.43
21	-0.67	0.88	-4.8	55.14	0.53
22	-0.37	1.03	1.1	48.23	0.29
23	0.09	0.99	-0.4	37.94	0.35
24	-0.52	0.97	-1.4	51.77	0.39
25	-0.49	0.93	-3.0	51.06	0.47
26	0.24	1.03	1.0	34.75	0.26
27	0.29	1.05	1.4	33.69	0.22
28	-0.33	1.01	0.4	47.34	0.31
29	0.07	0.98	-0.8	38.48	0.38
30	0.26	1.01	0.2	34.51	0.31
31	0.19	0.99	-0.3	35.93	0.35
32	-0.28	0.98	-0.9	46.19	0.38
33	0.31	1.01	0.4	33.45	0.29
34	-0.86	0.97	-0.9	59.40	0.38
35	-0.02	0.94	-2.1	40.50	0.45
36	-0.56	0.91	-3.7	52.75	0.50
37	0.44	1.03	0.7	30.73	0.28

Anmerkungen: Estimate = Itemschwierigkeit (aus Raschmodell), MNSQ = weighted mean square,  $T$  = Wert aus  $T$ -Verteilung,  $p_i$  = Itemschwierigkeit (nach Klassischer Testtheorie),  $r_{it}$  = Itemtrennschärfe

Tab. 1: Itemkennwerte

im Gesamttest vorhandenen Unterschiede hinausgehenden relativen Vorteil für Gymnasialschüler/innen und bei fünf Aufgaben einen Vorteil der Hauptschüler/innen (Differenz von mehr als 0.5 logits). In den Test werden perspektivisch nur solche Aufgaben aufgenommen, die keine oder nur geringe differenziellen Itemfunktionen aufweisen.

*Personenparameter:* Im Folgenden werden die im Gesamttest gemessenen Mittelwertunterschiede in den Personenfähigkeiten zwischen Schüler/innen als auch zwischen den drei Schulformen berichtet. Diese Analysen werden auf Basis derjenigen Items ohne differenzielle Itemfunktionen berechnet. Ziel ist es zu überprüfen, ob der Test die theoretisch anzunehmenden und/oder in der Literatur berichteten Unterschiede für den Bereich der Lesekompetenz abbildet. So sollten signifikante Schulformunterschiede sowie höhere Testleistungen bei den Mädchen bestehen.

Hierzu wurden für 566 Schüler/innen aller Schulformen zunächst individuelle Personenparameter (Weighted Likelihood Estimates, WLEs) mit der Software ConQuest erzeugt. Um die erhaltenen Werte besser interpretieren zu können, wurden die WLEs anschließend auf eine *T*-Wert-Skala ( $M = 50$ ,  $SD = 10$ ) transformiert. Die entsprechenden deskriptiven Daten für die Substichproben sind in Tabelle 2 aufgeführt.

Faktoren	Faktorstufen	<i>N</i>	<i>M</i>	<i>SD</i>
Geschlecht	männlich	266	52.35	7.55
	weiblich	300	51.59	6.93
Schulform	Hauptschule	205	47.71	5.80
	Realschule	189	52.85	6.31
	Gymnasium	172	55.99	7.08

Anmerkungen: *N* = Stichprobenumfang, *M* = arithmetisches Mittel, *SD* = Standardabweichung

Tab. 2: Deskriptive Daten für Lesekompetenz nach Geschlecht und Schulform

Zur Überprüfung des Einflusses des Geschlechts bzw. der Schulform (und dem Interaktionsterm beider Faktoren) auf die individuelle Lesekompetenzausprägung wurde in einem ersten Schritt ein latentes Regressionsmodell in ConQuest geschätzt. Diese messfehlerkorrigierte Analyse zeigt folgendes Bild: Es findet sich ein signifikanter Einfluss des Faktors Schulform auf die Höhe der Lesekompetenz; eine Geschlechtsspezifität ist dagegen nicht nachweisbar. Die Interaktion der Faktoren Geschlecht und Schulform bleibt ebenfalls ohne statistische Bedeutung (Tabelle 3). Mädchen und Jungen zeigen vergleichbare Leistungen in ihrer Lesekompetenz. Die nach Schulform differenzierten Ergebnisse verdeutlichen die erwarteten Schulformunterschiede.

Da die latente Regressionsanalyse lediglich einen Globaltest zur Signifikanzabschätzung liefert und keinen multiplen Mittelwertvergleich post hoc anbietet, wurde in

	Latente Regression mit ConQuest			Univariate Varianzanalyse mit SPSS		
Faktoren	<i>B</i>	<i>SE</i>	<i>T</i>	<i>F</i>	<i>p</i>	$\eta^2$
Geschlecht	–.083	.109	–.761	2.296	.130	.004
Schulform	.294	.081	3.630	82.027	.000	.227
Geschlecht × Schulform	.084	.052	1.615	0.922	.398	.003

*Anmerkungen:* abhängige Variable = WLE (transformiert auf *T*-Wert-Skala); *B* = unstandardisierter Regressionskoeffizient, *SE* = Standardfehler, *T* = Wert aus *T*-Verteilung, *F* = Wert aus *F*-Verteilung, *p* = Signifikanz,  $\eta^2$  = Effektgröße Eta-Quadrat

Tab. 3: Mittelwertunterschiede nach Geschlecht, Schulform und Geschlecht × Schulform in latenter Regressionsanalyse und univariater Varianzanalyse

einem zweiten Schritt eine univariate Varianzanalyse in SPSS berechnet. Hier zeigt sich ein vergleichbares Bild (Tabelle 3) wie bei der latenten Modellierung. Zusätzlich lassen sich die Schulformunterschiede zugunsten höherer Schulformen statistisch absichern. Demnach liegen die Ergebnisse der Gymnasialschüler/innen nahezu eine Standardabweichung über den Leistungen der Hauptschüler/innen.

Die zuvor berichteten Analysen dienen als Ausgangsbasis für weitere Testentwicklungsschritte. Für den dynamischen Test werden jedoch ca. 110 Items benötigt, sodass die jetzige Anzahl an Items zu verdreifachen sein wird. Bereits vorhandene Aufgaben werden überarbeitet und weiter genutzt.

#### 4. Diskussion und Ausblick

Zur experimentellen Überprüfung der Wirksamkeit unterschiedlicher Hilfen und Feedbacks im Rahmen eines computerbasierten Lesekompetenztests wurden bereits Texte und Aufgaben konstruiert, die für zukünftige Projektarbeiten weiterhin genutzt werden sollen. Eine Itemanalyse konnte zeigen, dass die bestehenden Aufgaben einem eindimensionalen Lesekompetenzmodell genügen. DIF-Analysen erbrachten nur bei einem Item eine unzulässige Bevorteilung der Mädchen; in Bezug auf die Schulform wiesen allerdings einige Items DIF auf. Diese Items wurden für nachfolgende Berechnungen zur Überprüfung von Mittelwertunterschieden nicht verwendet. In der Gesamtestleistung treten keine geschlechtsspezifischen Unterschiede auf, was bei sonst verstärkt nachgewiesener Geschlechtsspezifität im sprachlichen Bereich (vgl. Beck/Klieme 2007) im Zuge der Testkonstruktion weiter analysiert werden muss. Erwartungsgemäß gelingt hingegen die Differenzierung der Leseleistung zwischen Schüler/innen aller drei Schulformen. Zusammenfassend kann festgehalten werden, dass der pilotierte Test zu diagnostischen Zwecken herangezogen werden kann, jedoch weiter optimiert werden muss. Zudem kann derzeit nicht der Anspruch auf deutschlandweit gültige Aussagen erfüllt



werden, da die berichteten Ergebnisse aus einer nicht repräsentativen bayerischen Schulpopulation stammen.

Übergeordnetes Ziel der kommenden Projektjahre wird die Etablierung eines dynamischen Lesekompetenztests sein. Dieser Test soll einen breiten Einsatz in unterschiedlichen Fähigkeitsbereichen finden. Geplant ist es, den Test bundesweit in den Klassenstufen 5 bis 8 an Grund-, Haupt-, Realschule und Gymnasium einzusetzen.

Nachdem die Erprobung der entwickelten Textverstehensindikatoren den Grundstein für die weitere Testentwicklung legte, werden die bestehenden Materialien optimiert und für den Einsatz an Fünft- bis Achtklässler/innen hinsichtlich der gestellten Anforderungen in den Randbereichen der Itemschwierigkeiten erweitert. Basierend auf den Ergebnissen des Experiments zur Wirksamkeit von Fördermaßnahmen werden auch für die neuen Texte und Aufgaben Feedbacks und Lern-/Denkhilfen erstellt. Die Validierung einer so erzeugten dynamischen Version des Lesekompetenztests im Vergleich zur herkömmlichen Statustestapplikation (ohne Feedbacks und Hilfen) bildet einen Hauptstrang der zukünftigen Arbeiten. Ein zweiter Hauptstrang in der Testentwicklung ergibt sich aus der Erarbeitung und Evaluierung einer adaptiv dynamischen Variante des gleichen Lesekompetenztests.

Neben der technischen Umsetzung bietet die psychometrische Modellierung der Testdaten zahlreiche Herausforderungen: Die Schwierigkeit bei der Skalierung der Testdaten entsteht durch die mögliche Veränderung der Lesekompetenz mit Hilfe der angebotenen Feedbacks und Hilfen. Die Abbildung dieses Veränderungsprozesses und die Erfassung einer Lernfähigkeit stellen zentrale Herausforderungen beim Einsatz eines dynamischen Tests dar und werden weitere Betätigungsfelder im Zuge der Testentwicklung darstellen (vgl. Dörfler/Golke/Artelt 2009).

Inwiefern sich ein dynamischer Lesekompetenztest als Instrument zur Erfassung von Förderbedarf eignet, ist Gegenstand nachgelagerter Untersuchungen mit dem noch zu validierenden Test. Von Interesse ist hierbei z.B., ob sich Personen identifizieren lassen, die in besonderer Weise von den angebotenen Feedbacks oder Hilfen profitieren (sog. Gainer). Diese Testpersonen sprächen demnach gut auf Fördermaßnahmen im Bereich des Textverstehens an, woraus sich spezifisches Förderpotenzial ableiten ließe. Somit wäre durch die Etablierung des dynamischen Lesekompetenztests ein Schritt zur Vereinbarkeit von Diagnostik und Intervention getan.

## Literatur

- Artelt, C./Baumert, J./Julius-McElvany, N. (2003): Selbstreguliertes Lernen: Motivation und Strategien in den Ländern der Bundesrepublik Deutschland. In: Baumert, J./Artelt, C./Klieme, E./Neubrand, N./Prenzel, M./Schiefele, U./Schneider, W./Tillmann, K.J./Weiß, M. (Hrsg.): PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland. Opfaden: Leske + Budrich.
- Bangert-Drowns, R.L./Kulik, C./Kulik, J.A./Morgan, M.T. (1991): The instructional effect of feedback in test-like events. In: Review of Educational Research 61, S. 213–238.
- Beck, B./Klieme, E. (Hrsg.): (2007). Sprachliche Kompetenzen. Konzepte und Messung. Weinheim: Beltz.

- Coté, N./Goldman, S.R. (1999): Building representations of informational text: Evidence from children's think-aloud protocols. In: van Oostendorp, H./Goldman, S.R. (Hrsg.): The construction of mental representations during reading. Mahwah, NJ: Erlbaum, S. 169–193.
- Dörfler, T./Golke, S./Artelt, C. (2009): Dynamic Assessment and its Potential for the Assessment of Reading Competence. In: *Studies in Educational Evaluation* 35, S. 77–82.
- Golke, S./Dörfler, T./Artelt, C. (2009): Effects of Accuracy Feedback during a Text Comprehension Test. In: *Educational and Child Psychology* 26, S. 30–39.
- Graesser, A.C./Singer, M./Trabasso, T. (1997): Constructing inferences during narrative text comprehension. In: *Psychological Review* 101, S. 371–395.
- Guthke, J./Beckmann, J.F./Wiedl, K.H. (2003): Dynamik im dynamischen Testen. In: *Psychologische Rundschau* 54, S. 225–232.
- Hattie, J./Timperley, H. (2007): The power of feedback. In: *Review of educational research* 77, S. 81–112.
- Kintsch, W. (1998): *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- Kluger, A.N./DeNisi, A. (1996): Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. In: *Psychological Bulletin* 119, S. 254–284.
- Lewis, C./Reiman, J. (1993): *Task-centered user interface design: A practical introduction*. Boulder, CO: University of Colorado.
- Narciss, S. (2006): *Informatives tutorielles Feedback*. Münster: Waxmann.
- Oakhill, J.V./Cain, K. (2004): The Development of Comprehension Skills. In: Nunes, T./Bryant, P. (Hrsg.): *Handbook of Childrens Literacy*. Dordrecht: Kluwer Academic Publishers, S. 155–180.
- Oakhill, J.V./Garnham, A. (1988): *Becoming a skilled reader*. Oxford: Blackwell.
- Richter, T./Christmann, U. (2002): Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In: Groeben, N./Hurrelmann, B. (Hrsg.): *Lesekompetenz: Bedingungen, Dimensionen, Funktionen*. Weinheim: Juventa, S. 25–58.
- Shute, V.J. (2008): Focus on Formative Feedback. In: *Review of Educational Research* 78, S. 153–189.
- Souvignier, E./Mokhesgerami, J. (2006): Using self-regulation as a framework for implementing strategy instruction to foster reading comprehension. In: *Learning and Instruction* 16, S. 57–71.
- Wygotski, L.S. (1964): *Denken und Sprechen*. Berlin: Akademie.
- Willis, G.B. (2004): *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks: Sage Publications.
- Wu, M./Adams, R./Wilson, M. (1997): *ConQuest: Generalized item response modelling software*. Melbourne: Australian Council for Educational Research.

### **Anschrift des Autors/der Autorinnen**

Dr. Tobias Dörfler, Otto-Friedrich-Universität Bamberg, Markusplatz 3, D-96045 Bamberg  
E-Mail: tobias.doerfler@uni-bamberg.de

Dipl.-Psych. Stefanie Golke, Otto-Friedrich-Universität Bamberg, Markusplatz 3,  
D-96045 Bamberg  
E-Mail: stefanie.golke@uni-bamberg.de

Prof. Dr. Cordula Artelt, Otto-Friedrich-Universität Bamberg, Markusplatz 3,  
D-96045 Bamberg  
E-Mail: cordula.artelt@uni-bamberg.de